Luminosity and Analysis Tools

Adam Lyon

7 February 2003

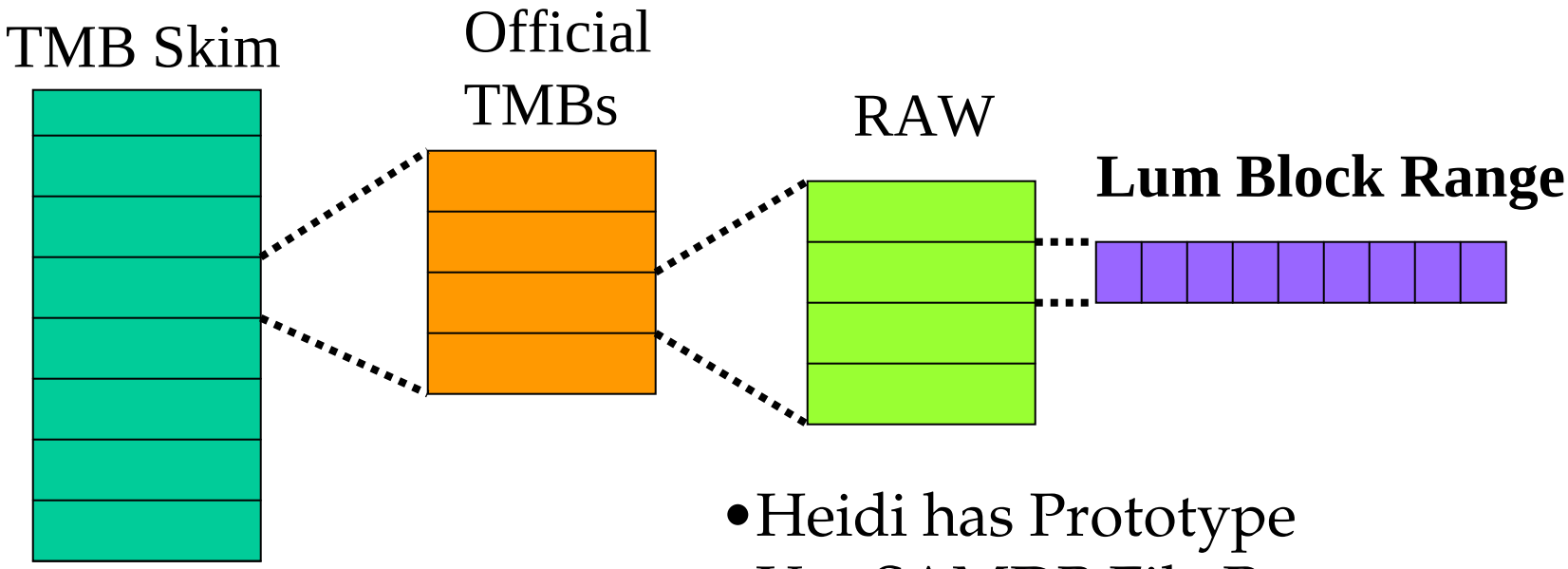
Luminosity Review

Breaking down the task

- Luminosity
  - Given a set of luminosity blocks, report the corresponding integrated luminosity
- Bookkeeping
  - Determine that set of luminosity blocks
  - Ensure that the analyzed data accurately correspond to the reported integrated luminosity
  - The *Analysis Tools Group* is helping with bookkeeping tasks
  - Interface with SAM, Runs, Trigger and Streaming databases
  - Perhaps use cases beyond just luminosity calculation

Bookkeeping Tasks

- Parent Luminosity Block Sample
  - Trace file parentage to the RAW file
  - Lum blocks are associated only with RAW files



- Heidi has Prototype
- Use SAMDB File Parentage
- Future direct hop to RAW

Ensure lum corresponds to analyzed data

- Did user actually analyze all the data they should have
  - SAM delivery failure (rare)
  - Analysis job crash
- SAM keeps track of files consumed for a project
  - SAM does the bookkeeping for you
  - But right now, there is no easy way to get this information (needs a tool)

Other sanity bookkeeping checks

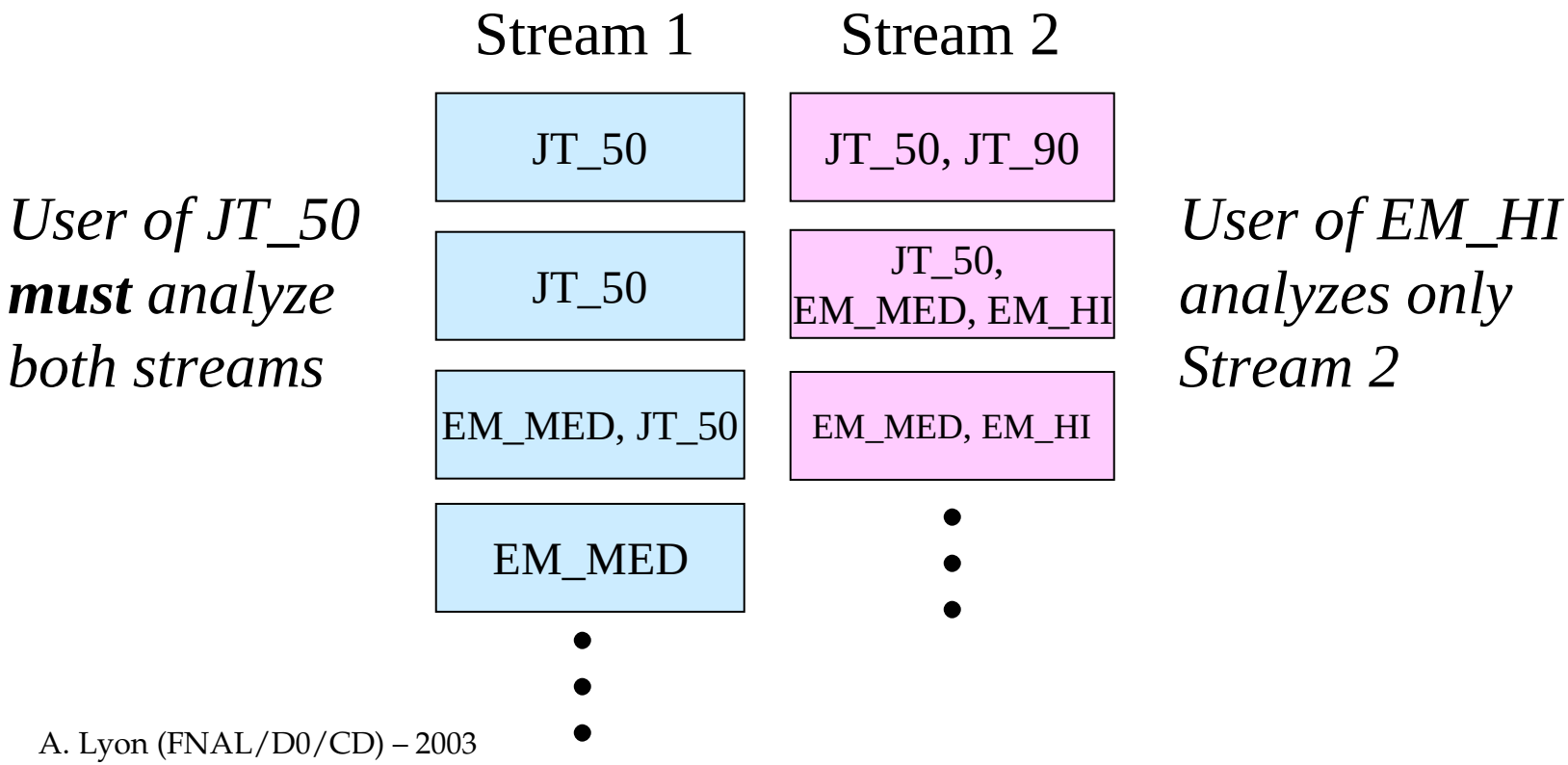
- Did user analyze duplicate events?
- Did events come from correct version of RECO?
- Were versions of RECO mixed?

How bulletproof does the system need to be?

My feeling is that most of these checks are nice, but should be lower priority than other bookkeeping tasks. Physicists must remain vigilant and keep track of what they do!

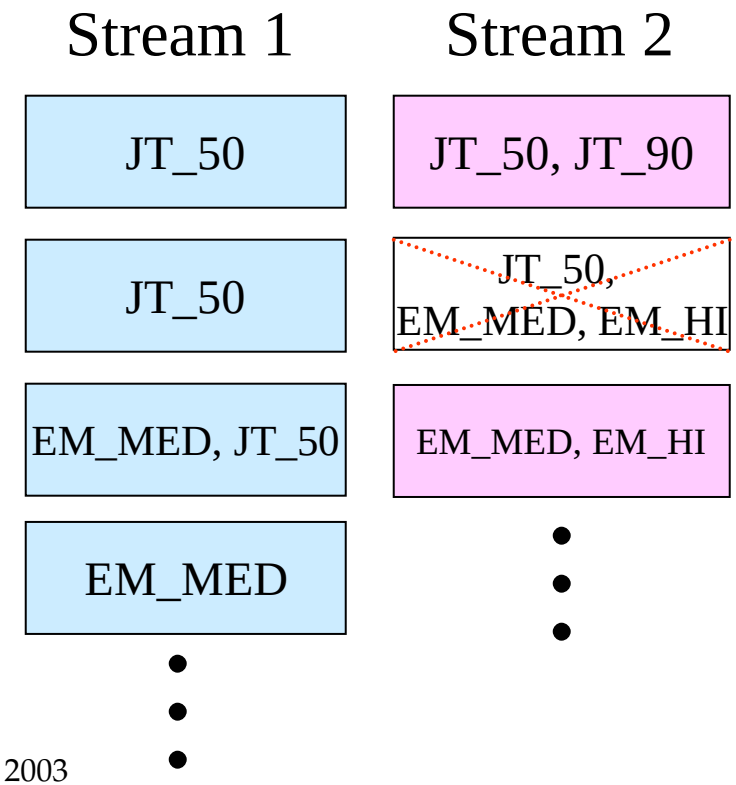
Streaming and bookkeeping

- Exclusive streaming:
  - Each event goes to **one** stream (events are not duplicated)
  - Events for your trigger may be in multiple streams



Streaming and bookkeeping

- Exclusive streaming:
  - Each event goes to **one** stream (events are not duplicated)
  - Events for your trigger may be in multiple streams



- What if user doesn't analyze a file for a stream?
  - If trigger can only go to that stream (EM_HI), then just a luminosity loss
  - If trigger can go to multiple streams (JT_50) then a biased loss

How worrisome is this?

- Farm losses
  - Unrecoverable RECO failures are tiny (~0.5%)
  - There are recoverable RECO failures (few %, <1% never recovered)
  - SAM store failures are small (few %)
    - Occur when storing files during SAM station problems
    - Recoverable (but haven't recovered any yet)
  - Merge fails only very rarely

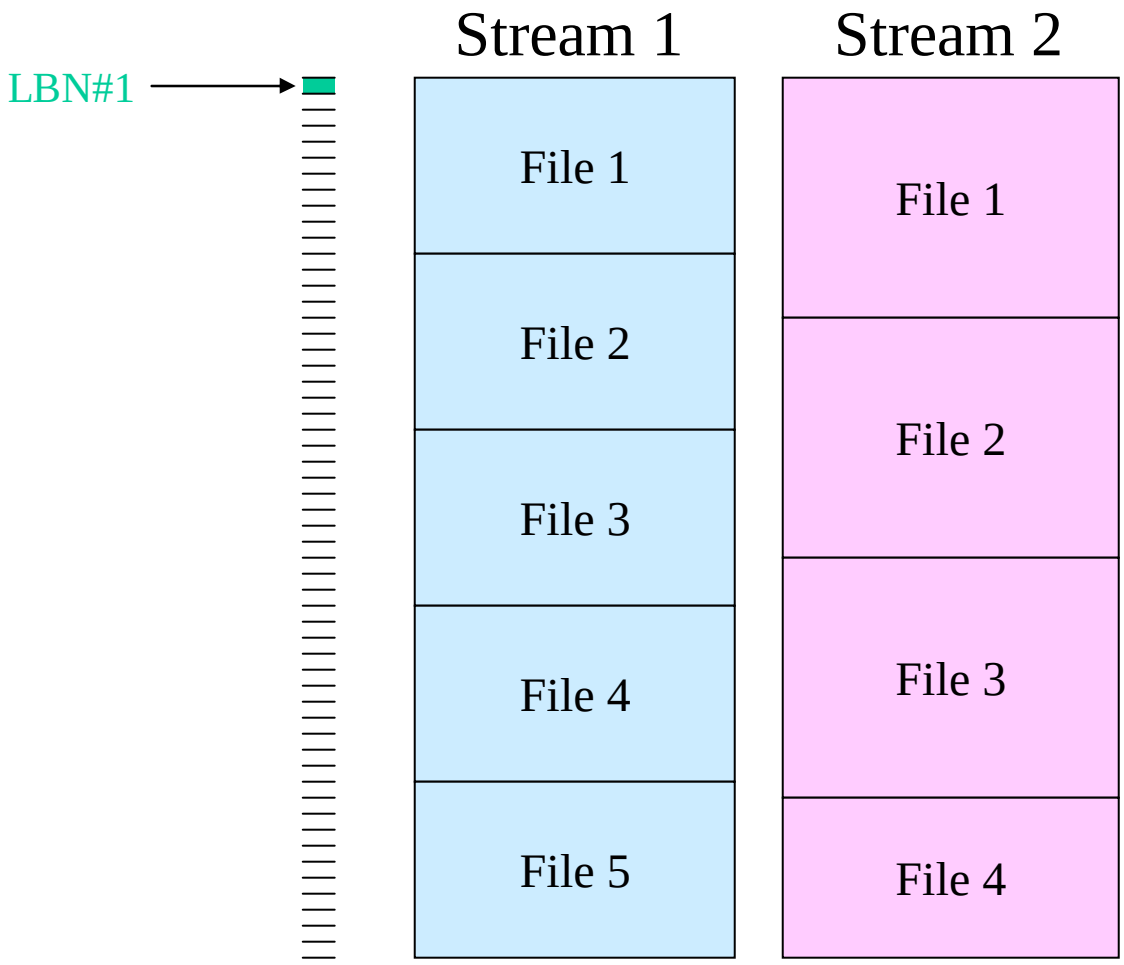
So farm loss rate is very small!

10% loss is a **worst** case scenario

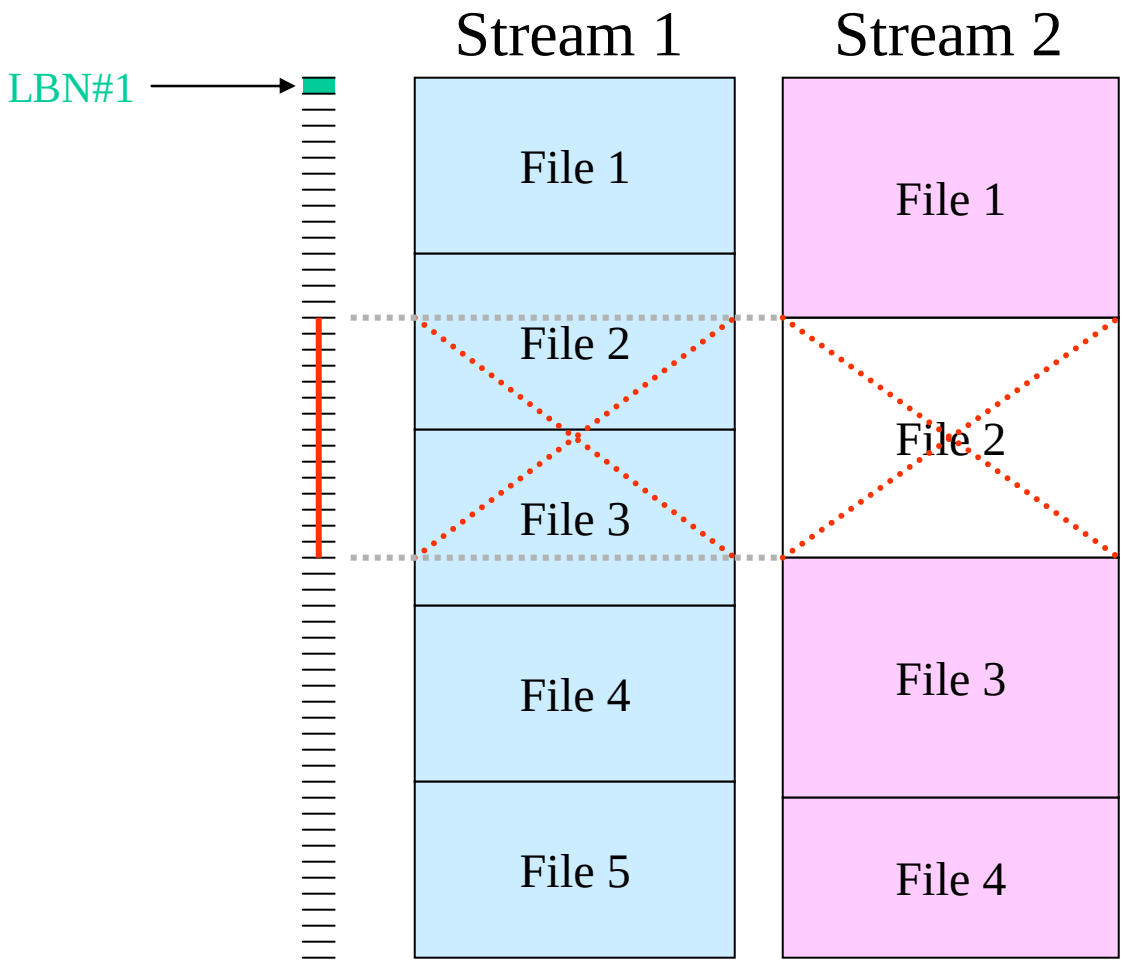
Effects of a biased loss

- Could be bad if you don't know about the loss:
  - Potential luminosity inaccuracy for precision analyses
  - Potential inaccuracy introduced to data-based efficiency studies
- If you know about the loss:
  - You remove events in your sample that came in the same luminosity block as the loss – this fixes the bias
  - Only lasting effect is decrease in integrated luminosity

Fixing a biased loss



Fixing a biased loss



Remove events from your analysis that are in the same luminosity blocks as the loss

Biased loss is rectified

Only integrated luminosity loss remains

Biased losses

- Most analyses won't care about a biased loss if it's small
- Obvious solution – provide a tool so the user knows about the loss!
  - 2nd bookkeeping tool to be written by ATG
  - Use file parentage and stream scheme to determine the streams that must be present for precise analysis
  - Checks that you consumed all the files you were supposed to analyze
  - User must record luminosity block # for every event in output root-tuple so certain blocks can be rejected

Other bookkeeping

- Bad runs
  - Remove bad runs that are in the dataset definition
  - Update luminosity without rerunning analysis job
  - Still needs to keep track of LBN# so can remove events newly tagged as bad from root-tuple
- Sanity checks
  - Consistent reco version (in future, better handled by SAM)
  - Duplicate events
- Framework to talk to luminosity system

The ATG Group and Priorities

- Manpower (as of now):
  - Me, Robert Illingsworth (Db Servers), consultations with Jeremy
- Past Deliveries:
  - Python tool to determine set of streams a trigger **could** have written events to (needed by online checking and offline tools)
- Top Priorities:
  - Biased loss catcher (involves parentage)
  - Bad run remover